

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Hillman, Jennifer L.  
Lal, Preeti  
Corley, Neil C.  
Shah, Purvi
- (ii) TITLE OF THE INVENTION: HUMAN CITRATE SYNTHASE HOMOLOG
- (iii) NUMBER OF SEQUENCES: 3
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.  
(B) STREET: 3174 Porter Drive  
(C) CITY: Palo Alto  
(D) STATE: CA  
(E) COUNTRY: USA  
(F) ZIP: 94304
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: To Be Assigned  
(B) FILING DATE: Herewith  
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Billings, Lucy J.  
(B) REGISTRATION NUMBER: 36,749  
(C) REFERENCE/DOCKET NUMBER: PF-0425 US
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: 650-855-0555  
(B) TELEFAX: 650-845-4166  
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 466 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: ENDCNOT03  
(B) CLONE: 2171653

Met 1	Ala	Leu	Leu	Thr 5	Ala	Ala	Ala	Arg	Leu 10	Leu	Gly	Thr	Lys	Asn 15	Ala
Ser	Cys	Leu	Val 20	Leu	Ala	Ala	Arg	His 25	Ala	Ser	Ala	Ser	Ser 30	Thr	Asn
Leu	Lys	Asp 35	Ile	Leu	Ala	Asp	Leu 40	Ile	Pro	Lys	Glu	Gln 45	Ala	Arg	Ile
Lys	Thr 50	Phe	Arg	Gln	Gln	His 55	Gly	Lys	Thr	Val	Val 60	Gly	Gln	Ile	Thr
Val 65	Asp	Met	Met	Tyr	Gly 70	Gly	Met	Arg	Gly	Met 75	Lys	Gly	Leu	Val	Tyr 80
Glu	Thr	Ser	Val	Leu 85	Asp	Pro	Asp	Glu	Gly 90	Ile	Arg	Phe	Arg	Gly 95	Phe
Ser	Ile	Pro	Glu 100	Cys	Gln	Lys	Leu	Leu 105	Pro	Lys	Ala	Lys	Gly 110	Gly	Glu
Glu	Pro	Leu 115	Pro	Glu	Gly	Leu	Phe 120	Trp	Leu	Leu	Val	Thr 125	Gly	His	Ile
Pro	Thr 130	Glu	Glu	Gln	Val	Ser 135	Trp	Leu	Ser	Lys	Glu 140	Trp	Ala	Lys	Arg
Ala 145	Ala	Leu	Pro	Ser	His 150	Val	Val	Thr	Met	Leu 155	Asp	Asn	Phe	Pro	Thr 160
Asn	Leu	His	Pro	Met 165	Ser	Gln	Leu	Ser	Ala 170	Ala	Val	Thr	Ala	Leu 175	Asn
Ser	Glu	Ser	Asn 180	Phe	Ala	Arg	Ala	Tyr 185	Ala	Gln	Gly	Ile	Ser 190	Arg	Thr
Lys	Tyr	Trp 195	Glu	Leu	Ile	Tyr	Glu 200	Asp	Ser	Met	Asp	Leu 205	Ile	Ala	Lys
Leu	Pro 210	Cys	Val	Ala	Ala	Lys 215	Ile	Tyr	Arg	Asn	Leu 220	Tyr	Arg	Glu	Gly
Ser 225	Gly	Ile	Gly	Ala 230	Ile	Asp	Ser	Asn	Leu	Asp 235	Trp	Ser	His	Asn	Phe 240
Thr	Asn	Met	Leu	Gly 245	Tyr	Thr	Asp	His	Gln 250	Phe	Thr	Glu	Leu 255	Thr	Arg
Leu	Tyr	Leu	Thr 260	Ile	His	Ser	Asp	His 265	Glu	Gly	Gly	Asn 270	Val	Ser	Ala
His	Thr	Ser 275	His	Leu	Val	Gly	Ser 280	Ala	Leu	Ser	Asp	Pro 285	Tyr	Leu	Ser
Phe	Ala	Ala	Ala	Met	Asn	Gly 295	Leu	Ala	Gly	Pro	Leu 300	His	Gly	Leu	Ala
Asn 305	Gln	Glu	Val	Leu 310	Val	Trp	Leu	Thr	Gln	Leu 315	Gln	Lys	Glu	Val	Gly 320
Lys	Asp	Val	Ser	Asp 325	Glu	Lys	Leu	Arg	Asp 330	Tyr	Ile	Trp	Asn 335	Thr	Leu
Asn	Ser	Gly	Arg 340	Val	Val	Pro	Gly	Tyr 345	Gly	His	Ala	Val	Leu 350	Arg	Lys
Thr	Asp	Pro 355	Arg	Tyr	Thr	Cys	Gln 360	Arg	Glu	Phe	Ala	Leu 365	Lys	His	Leu
Pro	Asn	Asp 370	Pro	Met	Phe	Lys 375	Leu	Val	Ala	Gln	Leu 380	Tyr	Lys	Ile	Val
Pro 385	Asn	Val	Leu	Leu 390	Glu	Gln	Gly	Lys	Ala	Lys 395	Asn	Pro	Trp	Pro	Asn 400
Val	Asp	Ala	His	Ser 405	Gly	Val	Leu	Leu	Gln 410	Tyr	Tyr	Gly	Met	Thr 415	Glu
Met	Asn	Tyr	Tyr 420	Thr	Val	Leu	Phe	Gly 425	Val	Ser	Arg	Ala	Leu 430	Gly	Val
Leu	Ala	Gln 435	Leu	Ile	Trp	Ser	Arg 440	Ala	Leu	Gly	Phe	Pro 445	Leu	Glu	Arg
Pro	Lys 450	Ser	Met	Ser	Thr	Glu 455	Gly	Leu	Met	Lys	Phe 460	Val	Asp	Ser	Lys

Ser Gly  
465

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: ENDCNOT03
- (B) CLONE: 2171653

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGCCGGTTTCG	TCTACTCTTT	CCTTCAGCCG	CCTCCTTTCA	ACCTTGTCAG	CCCGTCGGCG	60
CGGCCTCTGG	TGCAGCGGCG	GCGGCTCCTG	TTCCTGCGCG	AGCTCTCTCC	CTTTCTTACC	120
TCCCCACCAG	ATCCCCGAGA	TCGCCCCGCA	TGGCTTTACT	TACTGCGGCC	GCCCGGCTCT	180
TGGGAACCAA	GAATGCATCT	TGTCTTGTTT	TTGCAGCCCG	GCATGCCAGT	GCTTCCTCCA	240
CGAATTTGAA	AGACATATTG	GCTGACCTGA	TACCTAAGGA	GCAGGCCAGA	ATTAAGACTT	300
TCAGGCAGCA	ACATGGCAAG	ACGGTGGTGG	GCCAAATCAC	TGTGGACATG	ATGTATGGTG	360
GCATGAGAGG	CATGAAGGGA	TTGGTCTATG	AAACATCAGT	TCTTGATCCT	GATGAGGGCA	420
TCCGTTTCCG	AGGCTTTAGT	ATCCCTGAAT	GCCAGAAACT	GCTACCCAAG	GCTAAGGGTG	480
GGGAAGAACC	CCTGCCTGAG	GGCTTATTTT	GGCTGCTGGT	AACTGGACAT	ATCCCAACAG	540
AGGAACAGGT	ATCTTGGCTC	TCAAAAGAGT	GGGCAAAGAG	GGCAGCTCTG	CCTTCCCATG	600
TGGTCACCAT	GCTGGACAAC	TTTCCCACCA	ATCTACACCC	CATGTCTCAG	CTCAGTGCAG	660
CTGTTACAGC	CCTCAACAGT	GAAAGTAACT	TTGCCCCGAG	ATATGCACAG	GGTATCAGCC	720
GAACCAAGTA	CTGGGAGTTG	ATTTATGAAG	ACTCTATGGA	TCTAATCGCA	AAGCTACCTT	780
GTGTTGCAGC	AAAGATCTAC	CGAAATCTCT	ACAGAGAAGG	CAGCGGTATT	GGGGCCATTG	840
ACTCTAACCT	GGACTGGTCT	CACAATTTCA	CCAACATGTT	AGGCTATACT	GATCATCAGT	900
TCACTGAGCT	CACGCGCCTG	TACCTCACCA	TCCACAGTGA	CCATGAGGGT	GGCAATGTAA	960
GTGCCCATA	CAGCCATTTG	GTGGCGAGTG	CCCTTTCCGA	CCCTTACCTG	TCCTTTGCAG	1020
CAGCCATGAA	CGGGCTGGCA	GGGCTCTCC	ATGGACTGGC	AAATCAGGAA	GTGCTTGTCT	1080
GGCTAACACA	GCTGCAGAAG	GAAGTTGGCA	AAGATGTGTC	AGATGAGAAG	TTACGAGACT	1140
ACATCTGGAA	CACACTCAAC	TCAGGACGGG	TTGTTCCAGG	CTATGGCCAT	GCAGTACTAA	1200
GGAAGACTGA	TCCGCGATAT	ACCTGTCAGC	GAGAGTTTGC	TCTGAAACAC	CTGCCTAATG	1260
ACCCCATGTT	TAAGTTGGTT	GCTCAGCTGT	ACAAGATTGT	GCCCAATGTC	CTCTTAGAGC	1320
AGGGTAAAGC	CAAGAATCCT	TGGCCCCATG	TAGATGCTCA	CAGTGGGGTG	CTGCTCCAGT	1380
ATTATGGCAT	GACGGAGATG	AATTACTACA	CGGTCTGTGT	TGGGGTGTCA	CGAGCATTGG	1440
GTGTACTGGC	ACAGCTCATC	TGGAGCCGAG	CCTTAGGCTT	CCCTCTAGAA	AGGCCCAAGT	1500
CCATGAGCAC	AGAGGGTCTG	ATGAAGTTTG	TGGACTCTAA	GTCAGGGTAA	AACTGGAGAC	1560
TGGGTGAAAG	TGACTACCAG	AAAGTGAGGA	AGCCTAAATA	AAAAGTATAC	TTTTTGTTC	1620
GGGGGCCTTT	AAAGACTTAA	GATTAAATTA	TATCTGAGGC	ACTGATAATA	TGTTTGAGGT	1680
TAAAAATATA	ATTAAGACTT	TAAAAGATGA	AAAATGGTCC	CTTCTTCCCT	AATCAGCTCC	1740
CTTCCCCTGC	CTGGTATGAG	TTGCCCATCA	TACGCATGGT	CCTGGAGGAT	GACCAAGACT	1800
AATGCATGTG	GTATGAGTAG	GTTTGGCCCC	CTCACTATCT	CTAGAGTGAG	AATCTGGCTC	1860
CTGTTTCCAT	GGGTCAAAGC	CGGTTCGAGA	GAATCTGTAG	TCACTTTGGA	GCTTTAGCTT	1920
CTCTGCCAAG	CCCTCAATAA	GCCAGCAAAC	CAGGACTCTG	CCCCTTCTGT	TTCCATAGGA	1980
ATCATGTTGG	ATAGTCAGCT	GTACCAAGCC	CCTTGGCCCT	CTCCCATGCA	CACAAACACC	2040
TCCTAGCAAG	ACCTGTTGGT	TAGCTGGACA	TGCTTTGGCA	ATTTTTTTAT	ACTACCAAGT	2100
GACCATATTG	GCATGGCATT	TTTTGGTGAT	G			2131

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 464 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

(A) LIBRARY: GenBank

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

4



19-00000